DOCUMENT RESUME

ED 196 382 HE 013 380

AUTHOR Henderson, Euan; And Others
TITLE Developmental Testing for Credit: An Account of Open University Experience.
INSTITUTION Open Univ., Walton, Bletchley, Bucks (England). Inst. of Educational Technology.
PUB DATE Nov 80
NOTE 33p.
AVAILABLE FROM The Open University, Institute of Educational Technology, Walton Hall, Milton Keynes, England MK7 6AA.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Art Education; Case Studies; *College Curriculum; Cost Effectiveness; Credit Courses; *Curriculum Development; Curriculum Evaluation; *Curriculum Research; Earth Science; *Feedback; Foreign Countries; Higher Education; History Instruction; *Instructional Development; Music Education
IDENTIFIERS *Open University (Great Britain)

ABSTRACT

The experiences of course teams in developmental testing for credit between 1976-79 are described. The term "developmental testing," as used at the Open University, is defined as a stage in the course development process in which draft learning materials are tried out on students (or surrogate students) to obtain feedback before implementation. The place of developmental testing in the Open University's course development process is discussed, and developmental testing is compared with course improvement models that involve collecting feedback from students shortly after presentation. A rationale for the choice of testing for credit rather than testing for payment is also considered. Four case studies describing various aspects of the developmental testing of four courses (A241--elements of music, A292--Greece 478-336 BC, A101--arts foundation, and U201--risk) are presented, along with a note on the testing of another (S337--earth science). Each case study considers a number of issues, but each focuses primarily on a different aspect of developmental testing. Based on the case studies, the following conclusions about developmental testing for credit are offered: (1) this procedure does not, in general, lead to major structural revision of a course, and only incidentally to content changes, although it does result in a wide variety of modifications to the teaching of a course; (2) it can lead to benefits for students; (3) it can lead to benefits for authors; and (4) its costs and benefits are difficult to quantify, but the costs do not appear excessive. (Author/SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

IET

Institute of Educational Technology
The Open University
Walton Hall
Milton Keynes MK7 6AA
Telephone (0908) 65377

COURSE DEVELOPMENT GROUP

ED196382

DEVELOPMENTAL TESTING FOR CREDIT
AN ACCOUNT OF OPEN UNIVERSITY EXPERIENCE

Euan Henderson, Barbara Hodgson,
Adrian Kirkwood, Paul Lefrere,
Ellie Mace and Barbara Mayor

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ABSTRACT

This collection of papers describes course teams' experiences of developmental testing for credit between 1976 and 1979. During these four years, seven teams employed this approach to course improvement: A241 (1976), A101 (1977), A292 and E362 (1978), and D324, S337 and U201 (1979).

The Introduction discusses the place of developmental testing in the Open University's course development process, compares it with course improvement models which involve collecting feedback from students shortly after presentation, and provides a rationale for the choice of testing for credit rather than testing for payment.

There follow four case studies describing various aspects of the developmental testing of A241, A292, A101 and U201, and a note on the testing of S337. Each case study touches on a number of issues, but focusses primarily on a different aspect of developmental testing. Collectively, these case studies lead to the following conclusions about developmental testing for credit.

1. This procedure does not, in general, lead to major structural revision of a course, and only incidentally to content changes. It does, however, result in a wide variety of modifications to the teaching of a course (see especially the A241 and A292 case studies and the note on S337).

2. It can lead to benefits for the students (see especially the A241 and A292 case studies).

3. It can lead to benefits for authors (see especially the A101 case study).

4. Its costs and benefits are difficult to quantify but, insofar as quantification is possible, the costs do not appear excessive (see especially the U201 case study).

CD9 Developmental Testing/Course Planning/External Assessors/Critical Readers

Nov. 1980

1. INTRODUCTION

Adrian Kirkwood and Euan Henderson

The term developmental testing, as used at the Open University, describes a stage in the course development process in which draft learning materials are tried out on students (or surrogate students) in order to obtain feedback on the basis of which the materials may be revised before publication[1]. As such, it is only one of a number of strategies that course teams may adopt to improve a course during its development.

These strategies can conveniently be divided into two categories: critical commenting and student try-out. Critical commenting involves examination of how a course looks by a range of people who may be more or less expert in various aspects of the course. Student try-out is essentially concerned with finding out how a course works.

Critical commenting can involve a wide variety of people, including members of the course team and others (for example, external assessors and part-time tutorial staff). Different members of a course team may (and should) adopt different roles as commenters on draft materials; thus the perspectives of faculty academics, staff tutors, senior counsellors, IET staff, editors, etc. will often lead to the application of different criteria and expertise[2].

However, all of the activities within the broad category of critical commenting can be contrasted, in two important ways, with methods of trying a course out on students. First, critical commenting procedures tend to be more concerned with the content of a course, whereas student try-out tends to be more concerned with learning issues. Secondly, critical commenting tends to be particularly important in the early stages of course planning and in relation to earlier drafts of materials, whereas trying out the course on students comes towards the end of the course development process. Neither of these distinctions (content/teaching and early/late) are absolute, of course, but they do imply differences of emphasis between the goals of critical commenting and student try-out.

Alternative Models for Student Try-out

Developmental testing represents a particular model for course improvement through obtaining feedback from students. It usually occurs in the pre-presentation year, i.e. the best available drafts are tried out on a small group of students in the year immediately preceding that in which the course is to be presented for the first time to the undergraduate population at large. However, a number of other models have been, or are being, employed to improve courses by obtaining feedback from students shortly after publication.

In the 1+7 model (assuming a normal course life of eight years) a course is presented in the usual 'glossy' form and with all its components. During its first year of presentation, feedback is collected from some or all of the students and, if necessary, revisions are made and a new, or partially new, version of the course is published for the second and subsequent years of its life[3].

The 2+6 model (again assuming a normal course life of eight years) also involves presenting the course in the normal way, but it remains in unrevised form for two years. During the first year feedback is collected from some or all of the students and any necessary revision is undertaken during the second year. A new, or partially new, version is then presented for the third and subsequent years[4].

Rolling remake describes a number of variations on a theme rather than a single clear-cut model. It involves more or less continuous evaluation of a course, based on student feedback, throughout its life. Each year, certain sections or components of the course receive particular attention and are revised if necessary[5].

Each of these models has its inherent advantages and disadvantages, which may be summarised in terms of the following six issues.

1. Completeness of the materials. In the 1+7, 2+6 and rolling remake models, it is the complete course, in 'glossy' form with all components present, which is being evaluated. With developmental testing, materials are tried out in as near the final form as possible, but they will never be entirely complete: for example, the units will be reprographed rather than printed and some components may be missing entirely (notably, as a result of present production schedules, some or all of the television).

2. Size and representativeness of the student group. The 1+7, 2+6 and rolling remake models offer a full year's student enrolment from which to collect feedback whereas with developmental testing it has been usual to use a sample in the region of 25-40 (though in principle a larger number could be used). The use of larger samples is attractive, in the sense of ensuring representativeness, but there are disadvantages, most obviously the logistical problems associated with handling the correspondingly large amount of feedback data. With larger samples, it therefore may become necessary to collect largely closed-ended data, but it is open-ended data which is often more useful in drawing conclusions about necessary revisions[6]. When using the 1+7, 2+6 or rolling remake models smaller samples within the total course population may have to be selected for close study and the size and representativeness of the student group may not therefore be an important factor in selecting between models.

3. Effect upon students. With developmental testing, only a small group of students study the untried course, whereas with any of the other models at least one full cohort of students will study a course which has not yet been tried out on students. If there are serious concerns about the 'learnability' of the course, developmental testing has an advantage in this respect.

4. Time for making revisions. Developmental testing imposes greater pressure on the course team than the 2+6 and rolling remake models in the sense that the time available to collect and process feedback and to make revisions is shorter. The 1+7 model is generally agreed to impose even greater pressure on the course team in this respect.

5. Human resource implications. Any strategy for course improvement requires the input of additional time and effort by members of a course team. It does not seem to be the case that the total additional time and effort is significantly different between the four models of student try-out under discussion. However, they do differ in two important ways with respect to the human resource required. First, the total time and effort involved is spread over different periods: in developmental testing the extra resource is required in the pre-presentation year, in the 1+7 model it is spread over two years, in the 2+6 model over three years, and in rolling remake over most of the life of the course. Secondly, and arising from this, the extra work may fall on different people. For example,

in developmental testing the full load will be borne by the original production course team, whereas in the other models there may be some changes in the composition of the course team after presentation so that different people take on the work of evaluation and revision. Or again, in developmental testing there may be little or no extra editorial load (though the number of editor-days may be condensed into a shorter period), whereas in the other models any revision undertaken will require additional input from editors and possibly designers.

6. Costs. In addition to the extra human resources required by any of these approaches to course improvement, there are additional 'real money' costs. It is often very difficult, within the Open University, to attach even approximate 'real money' costs to various activities, and the activities under discussion here are no exception. However, it seems likely that with smaller population courses the 1+7, 2+6 and rolling remake models become rather expensive, since they involve, for example, short print runs. With larger population courses, on the other hand, they become (relatively) less expensive. What is not known is the relative costs of these three models, and the comparative costs between these and developmental testing. It may be, for example, that there is a break-even point, in terms of course population, between developmental testing and the 2+6 model.

Alternative Approaches to Developmental Testing

So far, developmental testing has been discussed as a single model, contrasting this approach to course improvement with those described as critical commenting and with other approaches involving student try-out. However, course teams at the Open University have engaged in a number of different activities which can properly be described as developmental testing.

Quite frequently developmental testing has been used in a limited, but important, way to try out small sections or components of courses. For example, a course team may wish to check that the first couple of units of a course 'lock on' to a recommended prerequisite. It therefore invites a few students who have just completed the prerequisite course to read the early units of the new course, in draft form, perhaps during the December-February period. Or a course team may wish to try out a home experiment kit, or certain experiments in a kit: surrogate students (perhaps Walton Hall staff or friends) are often used for this kind of testing. Sometimes these developmental exercises involve payment of the students involved; sometimes no payment is offered or required. This form of developmental testing, though small-scale and ad hoc, can be extremely valuable[7].

For other courses, developmental testing has involved try-out of the whole course on a group of students. If the maximum benefit is to be derived from this form of developmental testing, the course must be in as complete and final form as possible. Before reaching this stage, it is therefore important that such improvement as can be achieved by the various processes of critical commenting has been undertaken first, and it has generally been the case that prior to developmental testing the materials have passed through various draft stages, with the benefit of revision based on the comments of members of the course team, external assessors, and possibly part-time tutors. In fact, the form of the course which undergoes developmental testing should be, and has been, that which otherwise would have been handed over to the editor for publication (and sometimes may even incorporate some preliminary editorial and design work).

The demands made on students participating in developmental testing are considerable. Two methods have been used to compensate them for their efforts. One is to pay them (either in cash or, for example, in book tokens): this is generally described as developmental testing for payment. The other is to give them the opportunity, while they are studying the course in draft form and providing feedback, of submitting assignments, taking an end-of-course examination, and thus obtaining a credit towards their degree: this is known as developmental testing for credit.

Before 1976, only testing for payment had been used, and this method had three important disadvantages. First, it was associated with high drop-out rates (reaching 100% in some cases). Secondly, one component of the course rarely demanded of paid testers (and rarely tackled properly when it was demanded) was the assessment. Thirdly, the quality of feedback provided by student testers was often poor. These problems - high dropout, lack of reliable student performance data, and poor quality feedback - led to a reluctance on the part of course teams to revise materials in the light of developmental testing because of a loss in the credibility which could be attached to the feedback data.

In consequence, developmental testing for credit was introduced. The rationale for this was that payment was failing to secure the commitment from student testers. It was believed, however, that they would stay the course, complete assessment tasks, and take the matter of providing feedback seriously if they were motivated in the same way that many other Open University students are known to be motivated - by the opportunity of receiving a credit. Experience to date suggests that this judgement was correct: dropout rates during developmental testing for credit has been comparable with those normally experienced within the undergraduate programme generally, the great majority of student testers do submit assignments and take the examination, and the quantity and quality of feedback provided is generally much higher than in the case of developmental testing for payment. Though this seems to be largely attributable to the motivation provided by working for credit, it should be mentioned that these students also receive their set books free and, if they complete the feedback programme, are reimbursed their course fee.

Notes

1. The use of the term developmental testing at the Open University is somewhat idiosyncratic. For a full discussion of terminology in this field, see M. B. Nathenson and E. S. Henderson Using student feedback to improve learning materials pp. 25-32 (Croom Helm, London, 1980).

2. See: J. Riley Discussing and evaluating drafts (Course Development Paper No. 3, IET) and M. B. Nathenson and E. S. Henderson op. cit. pp. 35-9).

3. See: C. Newey TS251 An introduction to materials: "1+7" and its variants - an alternative to developmental testing (Courses Committee 32/15).

4. See: M. Nathenson The evaluation of T101: "2+6" model (IET Information Bank 1010).

5. See: E. Mace 'Rolling remake: an alternative approach to course design at the Open University' Programmed Learning and Educational Technology, Vol. 13, 1976, No. 4, pp. 25-30.

6. For further discussion of sample size see: M. B. Nathenson and E. S. Henderson op.cit. pp. 45-50.

7. For examples, see a memo. from Euan Henderson to the Course Resources Working Group on Developmental Testing, February 1980 (IET Information Bank 1232).

2. A241: A 'BETTER' COURSE FOR STUDENTS?

Euan Henderson

A241 Elements of Music was the first Open University course to undergo developmental testing for credit. In 1976, its pre-presentation year, all 16 of its unit texts were sent out, in sequence, to a group of London applicants who had failed to gain admission to the Arts Foundation Course that year. This group, initially of 28, worked through the course in reprographed, unedited, 'final draft' form, provided written feedback through the medium of 'feedback questions'[1], submitted assignments to their tutor in the usual way, and took an end-of-course examination. They were supplied with set books and a recorder free, were loaned a home experiment kit (a reed keyboard instrument), were not charged a course fee, and could gain a half-credit through the normal procedures of an examination board. Apart from receiving their printed material in reprographed, rather than printed form, the main difference in the course they studied lay in the broadcast material. Neither radio nor television could be produced by the BBC in time for the pilot run: unit authors made cassettes to simulate the planned radio programmes, but there was no analogue for the planned television programmes.

The feedback on each unit was processed by the educational technologist on the course team, who wrote a report for the author summarising the testers' views and experiences and, where appropriate, making concomitant recommendations for changes. All three authors took these reports very seriously and every unit was modified to some extent before handover for editing. In some, the changes called for and made were on a relatively small scale; in others, they were very substantial indeed - the most radical was to spread the teaching of the original unit 7 across two units (7 and 8) and to drop the planned final unit by way of compensation. In the case of the mock-up radio programmes, the use made of feedback from the pilot run was rather different: the experience provided to presenters was taken into account when rewriting scripts for BBC production.

The unit authors of A241 believed wholeheartedly that, as a result of using the feedback from testing, the final published version of the course represented a very significant improvement on what would have been printed if the 'final drafts' had been handed over, as they would otherwise have been[2]. They also believed that they had learned some general lessons which would be of value to them in writing for future courses:

> I learnt a great deal about the problems of the subject matter, especially the enormously varying degrees of difficulty encountered by [different] students ... I've learnt that concrete visual examples of points, which had seemed to be unnecessary, are often a very good idea, and that summings-up and checklists are always helpful. I had always underestimated the number of exercises that would be helpful. In future I'd offer more and suggest that people who had mastered material bypassed them[3].

> The main thing I think I learned was the need to be utterly clear and unambiguous in my writing; anything that stands any chance at all of being misunderstood is sure to be misunderstood by at least one student[3].

Another spin-off for the course team concerned its relationship with the operational areas of the University. Maintaining the developmental testing schedule was very hard work for authors (including the Course Team Chairman), IET staff, and course assistant alike, but the discipline of having to have units ready for the group of testers resulted in no slippage on this schedule and, most importantly from Operations' point of view, a remarkably close adherence to the handover schedule.

A third spin-off came in the tuition area. The tutor of the developmental testing group learned a great deal about the needs and problems in tutoring this particular course. A national tutor briefing was held at which, during the afternoon session, tutors were free to discuss the course with various members of the course team, as well as with the developmental testing tutor. It was noticeable that most of the tutors present spent most of their time in this experienced tutor's group.

Thus the course team believed that A241 was a better course as a result of the revision deriving from developmental testing, and that there had been valuable spin-offs from the exercise. But was the course actually better from the students' point of view? There are two difficulties in answering this question.

First, what does one mean by 'better' - by what criteria does one measure the improvement? This case study will examine four criteria: dropout, workload, performance (in continuous assessment and on the examination), and students' subjective views on, and experiences of the course.

Secondly, how does one examine these criteria in action - what experimental design will permit elimination of all extraneous variables? The short answer is that this is impossible, at least within present Open University constraints. The best design would perhaps be to divide the first year's enrolment into matched halves, with the two halves studying the pre- and post-developmental testing versions of the course, both in printed form. Even then there would be variables difficult or impossible to control. In A241, evidence was available from two sources.

(a) After the units had been revised in the light of feedback from developmental testing, the reprographed revised versions were studied by a second group of (initially) 20 testers. This group was also selected from unsuccessful London applicants, was matched as closely as possible with the first group (on educational background, musical background, and by means of pre-tests on the material in the earlier part of the course[4]), and worked under similar conditions towards a credit. This post-developmental testing group (referred to subsequently as 'post-DT') provided feedback in the same way as the pre-developmental testing group ('pre-DT') and the data from these two groups provides the most clear-cut evidence of the effects of changes made to the course as a result of developmental testing.

(b) Data is available on dropout and performance for the first two years of presentation of A241 (referred to subsequently as '1977' and '1978' groups). In addition, IET's Survey Research Department carried out a routine study on A241 in its first year of presentation and further data, especially on workload, is available from this[5]. It must be recognised, however, that the 1977 and 1978 groups

differed from the pre-DT and post-DT groups in a number of important ways, making direct comparison difficult: for example, they had already studied Open University courses, they were studying A241 in printed form, and television was included in their course.

Dropout

The dropout figures are given in Table 3.1. Direct comparison is impossible between the pre-DT and post-DT groups on the one hand and the 1977 and 1978 groups on the other: the latter were continuing students and the former were not, and it is known that dropout patterns differ considerably between these.

Table 3.1

	Dropout between final registration and the examination		
	1976	1977	1978
A241 pre-DT	50%		
A241 post-DT	50%		
A241		39%	31%
All Arts half-credits		32%	31%
All OU half-credits		29%	33%

Comparison of A241 with other half-credit courses is more illuminating. In 1978 the A241 dropout was average. The 1977 figure, however, was abnormally high. This is probably related to the rather curious mix of students who enrolled in that year. Student Progress Statistics show an enrolment with a very varied background, including substantial numbers whose previous Open University experience had been in other than Arts courses: only 66%, for example, had taken the Arts Foundation Course. This is very unusual in the Arts Faculty: for most courses the figure is 80-95%. It may be that a considerable number of students with some knowledge of, or interest in, music thought (mistakenly) that this course would provide an easy route to a half-credit. The enrolment pattern in 1978 was much less abnormal - partly, no doubt, due to information about A241 circulating informally amongst students and partly, perhaps, due to a change in the entry in the Courses Handbook.

There is no evidence that the revision of A241 as a result of developmental testing reduced dropout. But since we do not know how much dropout, University-wide, is course-related, we cannot say whether developmental testing should or could be expected to reduce dropout.

Workload

The mean unit study time reported by the pre-DT group was 12.6 hours, and by the post-DT group 10.8 hours[6]. The Survey Research Department report on the 1977 group indicated a mean unit study time of 10.6 hours. However, the pre-DT and post-DT figures are not necessarily directly comparable with the 1977 figure, since the ways collecting the data were different. Three points, nevertheless, can be made.

(a) A241 was at least two units overloaded before developmental testing. This overload was reduced by omitting the planned final unit of the course completely and by expanding a single earlier unit to make two, as well as by smaller reductions of workload across the whole course.

(b) A study of workload across the undergraduate programme conducted in 1976[7] indicated that only 4 of the 17 half-credit courses investigated had an average unit workload of 11 hours or less (based on Survey Research Department data); the other 13 averaged between 11.7 and 15.8 hours per unit (i.e. were overloaded by up to 7 units). Thus A241 is among the less overloaded half-credits, and the revision following developmental testing must take a large measure of the credit for this.

(c) Though the published version of A241 is not seriously overloaded as a whole, there remains a peak of workload at units 6-8 (11.7, 12.3, and 12.5 hours respectively according to 1977 data). This peak was significantly reduced as a result of developmental testing, but could not be eliminated entirely at this stage of production, even though the course team was made aware of it.

Performance

Performance measures are given in Table 3.2. The picture from this data is

Table 3.2

	Mean continuous assessment score	Examination pass rate	Mean examination score	Mean pass grade
pre-DT	67.3%	85.7%	59.6%	3.2
post-DT	78.8%	90.0%	63.4%	2.6
1977	75.1%	93.3%	66.3%	2.7
1978	81.0%	94.7%	62.8%	2.7

quite clear. On all four measures the pre-DT group were less successful than the post-DT group[6], and the performance in 1977 and 1978 closely reflected that of the post-DT group. The revisions arising from developmental testing did, therefore, lead to improved student performance in the course.

Students' Subjective Views

It may be argued that at least as important as the relatively objective measures of quality discussed above, and perhaps more important, are the subjective views of students about their learning experiences as they study a course. Feedback from the pre-DT group provided a wealth of detailed information on these students' learning experiences[6]. On the basis of this, many recommendations were made to the course team for changes, some major and some relatively minor, and the great majority of these changes had been made in the version studied by the post-DT group. This group responded to the same feedback questions, providing data which indicated, in very many cases, more positive subjective reactions.

Postscript[8]

The published version of A241 is not a perfect course - in particular some students have considerable difficulty with units 6-8 and this creates workload and tuition problems at this point in the course. One should not, however, expect perfection to be the outcome of developmental testing. What is clear is that developmental testing reduced workload, improved performance, and led to fewer problems for students. All members of the A241 course team agree, without reservation, that the course was tremendously improved as a result of the revisions made in the light of developmental testing. Some would say that, if these revisions had not been made before the course went out to a full cohort of students in 1977, extensive revision would have proved essential after publication.

Notes

1. For details of the techniques for collecting feedback used in the developmental testing of A241, see M. B. Nathenson and E. S. Henderson Using student feedback to improve learning materials Chapter 3 (Croom Helm, London, 1980).

2. They said so, quite forcibly, at a University seminar held on 10 May 1976.

3. Extracts from letters from course team authors to the present writer.

4. For an analysis of the two groups, see M. B. Nathenson and E. S. Henderson op. cit. pp. 172-3 or M. B. Nathenson, E. S. Henderson and B. K. Hodgson 'Developmental testing really does work', Improving Human Performance Quarterly, Vol. 6, 1977, No. 4, pp. 167-77.

5. P. Davey and C. Wareham A241 in 1977: Unit by unit feedback from students of Elements of Music (Survey Research Department, IET, 1978).

6. For full details of unit-by-unit study times, assignment and examination scores, and examples of the subjective reactions of students in the pre-DT and post-DT groups, see M. B. Nathenson and E. S. Henderson op. cit. pp. 173-81 or E. S. Henderson, B. K. Hodgson and M. B. Nathenson 'Developmental testing: The proof of the pudding', Teaching at a Distance, 10, 1977, pp. 77-82.

7. S. Blacklock Workload (Survey Research Department, IET, 1976).

8 Dinah Barsham, Lucy Carolan and Bill Strang (members of the A241 course team) have all endorsed this account of the developmental testing of A241.

3. A292: THE STUDENT'S PERSPECTIVE

Barbara Mayor

What can developmental testing contribute that the normal process of course team commenting and editing cannot?

Many members of course teams, and IET members in particular, consider it an important part of their role to comment on course materials from the point of view of the 'mythical' student. Without student feedback, however, one cannot be sure how realistic their (often varying) perceptions of potential student problems are, and the unit author may understandably be reluctant to act on any recommendations for change. The experience of genuine students can add considerable weight to the arguments of one's colleagues. It may, of course, also refute them, and either way the course gains. From my experience of the developmental testing of A292 Greece 478-336 BC I could quote many examples of each process at work.

First, two cases where the student responses confirmed an IET commenter's suspicions that some aspect of the text was not functioning as the author intended. At the first draft stage it had been suggested to a unit author that students with no musical background might have difficulty interpreting the musical notation in a unit on the Greek theatre. Similarly it had been suggested to another author that a detailed chemical explanation of the process of firing pottery might be unintelligible to the majority of students. Without student feedback, however, neither author had been convinced that any change to the text was necessary. When concrete evidence became available that students were experiencing difficulty with both these sections, the authors were happy to make appropriate changes. The section on chemistry now became optional and the musical notation was supplemented by a verbal explanation. The IET member of the course team could not have achieved this, however justified he felt the case to be; only the student comments were capable of convincing the authors that changes were necessary.

On the other hand, some of the anticipated difficulties simply did not occur. One section of a history unit, for example, seemed weak because of a lack of student activities. But when I asked developmental testing students how they felt about this absence of activities the answer was a unanimous "Perfectly okay!"

Thirdly, and of still greater significance, were the problems highlighted by developmental testing which the IET commenter had not anticipated at all and which would probably have remained undetected by the course team. A striking example was an in-text question in a literature unit which had not seemed to pose any particular problem but which, as a matter of routine, I had asked the students to answer on a feedback sheet. This provoked an outburst of complaints from the developmental testing students, most of whom felt unable even to attempt the question.

In general, a rich source of unforeseen reactions was the 'Any Other Comments' section at the end of each feedback sheet, which was extensively used by students to make original suggestions for the improvement of the materials. Thus, whether the developmental testing helped us to convey more convincingly what we knew already or whether it told us something entirely new, it amply justified itself both in our eyes and in those of the rest of the course team[1].

It could be argued that some of the points identified during a developmental testing exercise, such as a lack of clarity in expression, inconsistency in the use of terminology, or inaccuracy in cross-referencing, would in any case have been picked up at a later stage during the editorial process. In

the more obvious cases, I would agree. There were several instances, however, where even a close reading did not immediately reveal where the problem lay, and it was only students with a vested interest in complete understanding who were likely to tease out the ambiguities. An example from a philosophy unit may illustrate the point. The author's second draft contained what seemed on the face of it an unambiguous sentence:

> This brings out the importance of emphasizing feelings rather than actions when talking about the mean.

However, an in-text question, inserted for the purposes of developmental testing, revealed that this apparently straightforward question was being radically misinterpreted by the majority of students in the context of the argument being put forward: the whole argument had to be redrafted. There were similar instances of misunderstanding arising from hidden ambiguities in the use of the terms romanticism, idealism, realism, and naturalism in an art history unit, which again only emerged through answers to in-text questions and TMAs. I find it hard to imagine that either the editor or a non-assessed student would have identified such problems.

Assessment

I believe that it is the inclusion of assessment which makes developmental testing for credit a more valid exercise than developmental testing for payment. In either case, we are looking for committed students, but those with a credit at stake clearly stand to lose more if their understanding of the subject matter is not adequate for assessment purposes. The feedback available is also qualitatively different. The TMA scripts for A292, which were photocopied for us by Assignment Handling, were passed on to the appropriate unit authors to take into consideration, alongside the developmental testing reports, whilst revising their units. Several of them commented that these had been a rich source of information to them and an invaluable measure of the effectiveness (or otherwise) of their units as distance learning materials. Normally such feedback only becomes available to authors once the course has been printed and is no longer susceptible to easy change.

One particularly notable aspect of the developmental testing of A292 was the testing of a wide range of TMAs with a view to discarding the non-viable questions and 'banking' the rest for use in future years. Four or five questions were tried out for each TMA, distributed in such a way as to try to ensure that each question was attempted by students of varying abilities. Each question was attempted by at most three or four students, so that any conclusion about its viability was based on limited evidence. Nevertheless some clear indications did emerge: for example, some seemingly innocuous questions proved almost impossible to answer, whilst others obliged students to go beyond the course material. Such questions were discarded[2]. Allowance was of course made by the examination board for students who had been disproportionately disadvantaged in the questions allocated to them. The exercise was generally considered to have been a success, and the course team has come out of it with a bank of proven questions for at least half of the planned life of the course - so far as I know, a unique achievement.

The Creative Role of Developmental Testing

A recurrent critcism of developmental testing is that it fails to come up with constructive suggestions. In A292 this criticism was shown to be ill-founded on two counts - the contribution made by myself as evaluator, and that made by the students.

As evaluator, I devised a number of questions which were designed specifically to extend the material, for example:

> Were there any passages of the set book you needed help with where help was not given?
>
> What extra information would you have liked to have?
>
> Can you suggest any alternative way of presenting this section?
>
> Would you have preferred another format?
>
> What further questions would you ask yourself?

In one particular case where the draft unit did not make much active demand of the students I inserted, by way of experiment, a number of additional activities in the feedback sheets, sometimes to provide a means of checking the student's understanding and sometimes specifically to test the viability of the inserted question as an in-text question. This experiment was successful, for not only did a number of misunderstandings emerge from the answers given but also several of the questions were subsequently incorporated by the unit author as in-text questions in his final draft[3].

So far as creative initiatives from the students were concerned, I have already mentioned the positive and original contribution which all students could make to the course via the 'Any Other Comments' sections of the feedback sheets. I should like to focus here on some of the cases where individual students exerted a substantial influence on the way in which the text was subsequently revised. It will be convenient to handle these under the broad categories of errors, questions, suggestions, and objections.

1. Student errors. Although I have used the term 'errors' it would often be more appropriate to speak of 'divergent answers'. I am thinking not only of cases where students actually got an in-text question wrong (which would normally suggest that an improvement in the initial teaching or an expanded discussion was called for) but also, as often happens in the humanities, of cases where a student's answer was apparently just as valid as the author's and perhaps raised interesting new questions. An illustration of this comes from a literature unit, where students were asked to think about the image of the net in Aeschylus' Oresteia. A student's tentative comment:

> I wonder whether the idea of the Fates spinning a thread of life (see Eumenides 334) can also be tied in with the ideas of the nets and tapestries?

led to the inclusion of this point in the author's redrafted discussion. Several examples of a similar nature occurred in a history unit where the author's discussion of the in-text questions seemed to be at a tangent to the answers which students were actually giving and substantial re-writing was necessary to bridge the gap. Thus, in a small way, the developmental testing student becomes 'co-author' of the text.

2. Student questions. There were numerous cases where individual students requested additional information on the issues raised in the text and authors were often able to work this into their redrafted texts. A brief reference to 'metics', for example, provoked one student to ask if the children of immigrants were considered aliens. A footnote was added briefly describing Athenian citizenship law. Similarly, in a literature unit, a student queried the colour of the royal carpet:

> Was the carpet purple or red? Are these colours distinguishable in Greek?

The text was subsequently amended to refer to the colour as 'purple or crimson', reflecting an ambiguity in the translation. There were many examples of a similar kind. In this way a little of the richness of face-to-face teaching was incorporated into the printed material.

3. Student suggestions. There were many cases where students commented on the pedagogy of the unit. Some suggestions were for a re-sequencing of the material on the basis of the student's experience of working through it. As a result of this, in at least two cases, set readings were recommended before starting work on the unit proper. In other cases, students suggested the inclusion of a visual aid such as a table or diagram - a problem which was solved on more than one occasion by referring them to an existing figure elsewhere in the course which the author had overlooked. In all these ways the teaching was made more student-centred. But the most memorable example of all was a case where the majority of students had either misinterpreted or failed to understand a particular in-text question, and a student suggested an alternative phrasing which eliminated the ambiguity. Her re-wording was adopted by the author almost exactly as it stood.

4. Student objections. In this category I am including several instances where authors wrongly anticipated a particular reaction or behaviour on the part of students. Students objected to the use of phrases such as 'hard as it may be to believe' and 'you wouldn't have noticed' when they did not in practice apply to the majority of them. Thus authors were sensitised to the students' actual performance and were able to adjust their teaching in response to it. Objections were also raised to signs of bias on the part of authors, as a result of which phrases such as 'the Puritan ethos of modern industrial and commercial organisations' and 'Fascist pig' were removed from the final version. Lastly, it took a woman to point out to a male author that his phrase 'the pains of childbirth before the days of anaesthetics' underestimated what many mothers still go through!

Developmental testing, by permitting communication like this, is the nearest approach which the Open University production process can provide to the living interaction between teachers and students out of which the whole learning process is created. One of the unit authors on A292 summed up his experience as follows:

> There [was] a certain amount of obligation on me when writing to look to those individuals and to see the comments that came back. It was a very much more personal thing. It was more like conventional teaching.

Costs and Benefits to the Student

What of the relative costs and benefits to the student involved in developmental testing? There is clearly some inconvenience involved:

> I found the typescript rather dull compared with the usual, beautifully produced Open University units. It was sometimes difficult to concentrate while reading from a mass of papers, stapled together. I was afraid of losing the feedback sheets, which were interspersed with sections of the unit ... The illustrations to the arts units came in packages of numbered plates ... A great deal of time was wasted in sorting and shuffling[4].

On the other hand, it has been found that students are far less reluctant to criticise a typewritten unit than the final printed version:

> I think it may well be that the actual finished product is too finished to stimulate any feedback. It may put people off. In the glossy final product I think I would have been somewhat scared of actually writing some of the things which one did write about parts of the draft material[5].

The problem of the inserted feedback sheets has subsequently been solved by other course teams, who have reprographed them separately from the units. There does not seem to be any way round the problem of providing illustrations of photographic quality, but good photocopies have since been found to suffice, except in certain disciplines such as art history.

On the benefit side, participation in the developmental testing exercise seems to have acted as a stimulus to study:

> I found it interesting to know, however, for the first time, just how long I spent on each unit ... I almost always completed each feedback sheet as I came to it. Far from being an interruption to study, I found it both a welcome diversion and a mark of progress ... [It] must have improved my understanding and absorption of the material ... I found it difficult to know whether I would have done the in-text questions if I had been a normal student so I went ahead and did them all. This enabled me to indicate when a question was valueless and to judge whether the number of questions was about right[4].

Many other students said informally that the experience of developmental testing had helped them to develop a critical perspective which would be an advantage in subsequent Open University study.

Most importantly, many developmental testing students said afterwards that they felt privileged to have been involved in the exercise:

> I am glad ... to have had the experience of developmental testing and seeing a course in the making. As students we were made to feel a valuable part of a team[4].

> [It] made us feel very much part of the course-building process.[5]

By including students and tutor on the production team in this way, developmental testing goes a long way towards creating course presentation conditions. This is the only way in which a course can and should be judged.

Notes

1. See also: L. Hardwick et al. 'So what were we testing?', Teaching at a Distance, 17, 1980, pp. 52-66. Lorna Hardwick, Course Team Chairman of A292, has endorsed this account of the developmental testing of A292.

2. For further analysis of this aspect of the testing, see the contribution of the developmental testing tutor, Eileen Powell, in L. Hardwick et al., op. cit.

3. I have developed this argument further in L. Hardwick et al., op. cit.

4. These quotations are taken from the contribution by June Glover, one of the A292 developmental testing students, to L. Hardwick et al., op. cit.

5. Comments by an A101 developmental testing student at a Production Methods Committee seminar held in February 1979.

4. A101: DEVELOPMENTAL TESTING AND THE COURSE WRITER

Ellie Mace

On the face of it one might wonder why it is necessary to write a paper in support of such an obviously sensible idea as trying out a course on its recipients before it is produced in a polished, and more or less permanent, form. The answer must be that there are costs involved (some obvious and some not so obvious) that are seen as outweighing the possible benefits. The more obvious costs are money, time and commitment. Developmental testing involves extra time and effort on the part of students, course writers and evaluator(s). Moreover, it requires that course writers commit themselves to the course (and thereby pass up other opportunities) for a longer period than normal. Because it means testing every aspect of a course, developmental testing also makes demands on administrative, BBC and editorial course team members. It may require all these people to change their usual ways of working and their schedules. It also places most course team members in different working relationships to one another and to the students. The evaluators (normally IET members of the team) become central to its organisation during this period. They set up the project, organise its execution, interpret the feedback between students and writers, and they are responsible for negotiating with writers' changes to their texts. For their part, writers are faced with the possibility of having to make extensive revision to their work at the point when, under normal circumstances, they would be handing it over for production. They may also be rather thrown by the kind of criticisms that students make and find that, at this late stage, they have to consider their work in a somewhat different light. In view of all this I think it is easy to see why many people have to be convinced that what they are going to gain from the experience is worth the trouble. In this paper I shall argue that the benefits are indeed on a scale such that pressure is put on those who oppose the idea of developmental testing to show how the resources could be better spent. In the main I shall discuss the matter from the point of view of the full-time Open University course writer, whose attitude towards course development and whose role in developmental testing is, I think, crucial. I shall argue that for the course writer developmental testing has certain advantages over other methods of course improvement (with reference, in particular, to the testing of A101 An Arts Foundation Course).

But first it is necessary to establish that developmental testing is bona fide a means of course improvement. Certain criticisms that have been made would appear to leave it without a theoretical leg to stand on. The first main criticism is that developmental testing can only ever result in improvements to parts of a course and not to the course 'as a whole': if so, in what sense (if at all) can we claim to achieve 'course' improvement by these means? The other concerns 'evidence' of such improvement. It is by no means clear how we can reach agreement that a course has, or has not, been improved by developmental testing; what then, is to count as 'evidence' of course improvement? I shall begin by taking these two issues in turn.

Developmental Testing and 'Course' Improvement

Most critics accept that developmental testing can provide course producers with certain sorts of information: about the efficacy of the teaching methods they have employed; the clarity with which they have presented discrete sections of the course material; how students feel about the course; how they tackle the material; and so forth. They also concede that such information can provide the basis for alterations, even improvements, to aspects or parts of a course. But they argue that developmental testing cannot reveal possible weaknesses in the

way a course is structured: even if students are able to yield the relevant information such testing cannot affect basic decisions, taken at a much earlier stage, regarding the overall strategy and design of a course (such decisions as what is to be taught; how it is to be taught - text or television, for example; the order in which it is to be taught). So it is perfectly possible for a course to be tested and extensively revised as a result, and yet for such a course to 'fail' (the whole being more than just the sum of its parts) because structural weaknesses persist. The conclusion is that developmental testing is not so much a means of improving a course, but more of tinkering with its component parts.

The force of this criticism rests on the assumption that the main problems with a course often arise out of the way it is structured: because developmental testing cannot get at the structure of a course, or at least can only fiddle with it, it is seen essentially as cosmetic. The assumption is arguable. Clearly there are many possible ways of structuring a course of study (and of combining courses and disciplines) so that taken at face value such a view smacks of Utopian thinking (i.e. there is a best way of structuring a course if only we can find it). Furthermore, in my experience it is rarely, if ever, the case that a course flounders because it is fundamentally ill-conceived. The fact that our courses are planned by course teams, involving a number of people many of whom are now experienced Open University producers, seems to make it unlikely - neither carelessness nor eccentricity flourishes in such a setting. However, I accept that the way a course is structured can cause problems for students, that the students' comments do not often reflect such difficulties and that, even if they did, by that stage in course production there might be little that could be done about it. But if those who write off developmental testing on such grounds overestimate the extent to which students' problems arise out of course structure, they also undervalue the importance and quality of the insights students can provide about course material (a point I shall explore more fully later on). To do so seems a bit like blaming the apple tree for not bearing plums.

The strong, or 'political', version of this argument, though, amounts to an attack on the way most Open University courses are designed, i.e. developmental testing merely reinforces the status quo by making more palatable courses that have been conceived, planned and executed largely by central academics for students, with little or no account taken of the students' purposes or interests. The real issue here seems to be whether the way courses are produced is sufficiently sensitive to the needs of students. The criticism does not, I think, undermine the purposes or techniques of developmental testing itself. Though developmental testing may not go far enough for such critics it is, ironically, the only systematic attempt so far to make course writers aware of the course material as perceived by the students. This potential for understanding would indeed be wasted if the application of developmental testing were to become a matter of routine, and its techniques a sort of formula - ways of confirming the fitness of our procedures and products, and of bringing people into line, rather than of gaining insights into the way they function. As a set of techniques, developmental testing is as 'good' as the people who operate it. When used sensitively and imaginatively it is still not a panacea. But, far from shoring up current practices, it is possible for the understandings gained from the testing of one course to be fed into the early planning stages of the next, and so indirectly to enable us to affect course structure. In future, if such critics can find a way of setting up courses of the kind they favour there is every reason to suppose that current developmental testing techniques could be adapted to yield the kind of information and insight they too would find useful and illuminating.

In the meantime developmental testing has had a more limited goal. Briefly, it assumes that the students' view of course material is unique (you cannot, for example, get access to it by asking tutors what they think it is), and that with the students' help we can construct courses that are both more accessible and more interesting (some might say more 'effective'). This does not preclude the possibility of the course team trying to improve the ways in which it reaches decisions about course structure or finding ways of improving structures that appear to be weak - and, as I said, even of taking account of students' ideas on the matter. But whether students have the right to participate in course planning, or whether out of respect for them we ought to consult them, are different and altogether trickier questions that I can only touch on here. My own view is that students, particularly early on in their careers, can only have a limited role to play with respect to course planning. Deciding what is worth including in a course of study, and how best to structure it, depends on thorough knowledge of the subjects of that study. In theory students are 'learners' and cannot be said already to 'know'. In practice, even mature or experienced students are not in a position to appreciate the relationship of a given course to its parts until they have at least studied the whole course and have had time to reflect on it. Even then their perspectives and judgements will change, and continue to change, in the light of further study and wider knowledge. I cannot imagine how such students could construct courses for themselves[1]. For these reasons it seems to me a mistake to expect students to contribute fully to discussion about course structure. But it is equally mistaken to think that because they cannot, they have nothing else to say of much value. Correspondingly, the fact that developmental testing does not throw much light on course structure does not make it the equivalent of a face-lift. In my view 'course improvement' properly describes the successful outcome of a systematic attempt to make a given course more accessible to as wide a range of students as possible.

'Evidence' of Course Improvement

The question now is what do we mean by success in this context: what is to count as 'evidence' of improvement? In the first place we can look to students for evidence of the success or failure of revisions made to courses on the basis of developmental testing. We can look at what they actually do as they work through the course, both before and after developmental testing, and try to compare their performances on these occasions[2]. We can also take into account what they say about how difficult they find the work, how long it takes them, how interesting they find it, and so on[3]. Here, and in the rest of this paper, I shall suggest that we can also look at the opinions of course writers. In what follows I shall argue that their judgements are just as valid as students', if different, and just as important as 'evidence' of improvement.

Immediately it might be objected that course writers are too close to their material to be able to make reliable or 'objective' judgements about it: they have jointly conjured up the course and fashioned its structure, as well as filling it out in detail - work that is difficult, personally demanding, and time-consuming. By the time the course undergoes developmental testing they have a strong intellectual and emotional commitment to it which, apart from making them reluctant to agree to the recommended changes, means that their judgements regarding the effectiveness of such changes are suspect. They are the worst possible people to ask.

I appreciate the force of this argument, but it seems to me to be mistaken. I think it is true that, for all the reasons outlined above, course writers often are understandably reluctant to make alterations to their work. But

if writers can be persuaded to make the recommended changes I would say that, just because they are committed to their original work, their subsequent comments should be taken seriously. We must assume that course writers, of all people, are aware of their own short- and long-term intentions, so if they are convinced that alterations suggested by developmental testing actually amount to improvements on what they have produced then we must surely take them at their word. What they have to say may not amount to 'hard data' but it is certainly 'witting testimony'.

The testimony of A101 writers does, I think, demonstrate that from their point of view the course was improved as a result of developmental testing[4-9]. All writers made most of the changes that were recommended to them which, for some, meant extensive revision to their work[7]. And it is interesting to note that some said they had learned general lessons from the experience that they felt would affect the way they approached their work in future[4,5]. Moreover, the feedback from developmental testing contributed to a course team decision about the structure of the course, i.e. to rewrite one unit and entirely replace another during the following year. One writer even found in students' work material for his own unit[9].

Course Writers and Developmental Testing

From the point of view of writers the main advantage of developmental testing is that it enables them to reach out to their audience. It is an occasion on which they can enter into the kind of negotiation that can take place regularly between teachers and students at conventional universities. Such teachers also make assumptions, based on experience and intuition, about the content of what they teach, the methods they use, the pace at which they proceed and so on. But they have the obvious advantage of being able to respond to what is going on - to change direction mid-stream, supplement a point, toss in an enlivening aside. By means of developmental testing Open University writers, at least on one occasion, have real students to address who can report back to them real problems and difficulties they have encountered at a time when those writers are in a position to do something to improve matters. The system is flexible enough to allow problems of both general and very specific kinds to emerge, and it allows students to make what seem to them to be realistic suggestions about how they might be overcome. I argued earlier that only students can provide this perspective. But I did not, and do not now, mean to suggest that, therefore, all courses should be tested in this way. Rather I think that if all course writers, particularly those who are new to the University, could experience developmental testing just once much would be gained. There is plenty of evidence that A101 writers, for example, found the process itself interesting and fruitful. Some saw it as an opportunity to experiment with ideas or teaching strategies that interested them but which, under normal circumstances, they might not risk (for example one writer wanted to see whether she could get across a number of points of view on a question by presenting them in the form of a 'dialogue' between people with conflicting ideas). Others were sufficiently interested in the outcome of the testing to examine the data in its 'raw' form[6] (the writer actually asked to see all the data), or to visit the students some considerable distance away in order to discuss their difficulties with them. And as we saw[4,5], some writers claimed that what they had learned from the experience would stay with them.

One can only speculate about why such exposure to students should have the impact that it does. After all, most course writers constantly meet students for a full week at summer schools. It seems to me that there are important differences between the two situations. The summer school setting is not one that encourages students to make critical comments in detail to course writers

about their work, whereas this is precisely the role of developmental testing students. Also, at summer school writers are not in the frame of mind to receive such criticism or in the position to put it to good use. When course writers 'meet' developmental testing students, however, they are actually involved in the process of writing for them. The attention of both writers and students is focussed on quite specific, and immediate, issues. And the writers really have to think hard about the nature and possible causes of students' difficulties in order to make appropriate revisions to their work. Under these circumstances they are perhaps able to form a stronger impression of how the students actually tackle their work and what their particular problems are than they are at summer school. And they can perhaps form sufficiently strong mental images of these real students to be able to recall them to mind when they embark on subsequent projects. In other words, as a result they write for 'real' students, rather than imaginary ones or 'models' of students: developmental testing perhaps helps these writers towards a more realistic sense of audience.

Another important reason for the impact of developmental testing on the writer may be that writers find they can actually use the information and insights it provides. As I have said, there is no formula for developmental testing. Writers and evaluators can collaborate to frame questions they hope will reveal the kinds of difficulties students may find in a particular unit, or particular section of a unit. Therefore the kinds of questions that predominate will be quite different from one unit to another[10]. One writer may want to try out a section of the text in the belief that students will find it conceptually difficult. A brace of questions of the type 'What main point(s) do you think the author is making in this section of the text?' and 'Do you feel you understand what he/she is saying?', followed by an answer to the first question and a further question of the kind 'If you got this wrong what do you think was the trouble (what caused it/what might be done about it)?' would be appropriate here. The answer to the first question should reveal whether most students have actually grasped the point. By comparing the first and second answers it should be possible to tell how confident they are feeling (students might, of course, have grasped the point but feel that they have not, or vice versa - and if so, we might suspect the writer of some ambiguity). The answers to the third question should provide clues to the nature of students' problems and suggest ways in which the text may be altered. Another writer, who for example relies on exercises to get the point across, will want to be sure that particular exercises function as they are intended to. In this case students would be asked to do the exercise, and afterwards to comment on what they thought its purpose was, how they fared, and so on. By looking at what the students actually do, and by reading their comments on it (and to some extent reading between the lines), the writer should be able to spot weaknesses in the design of the exercise and have some idea about how to improve it - or conclude that in this case it would be better to dispense with it altogether and explain the point in a straightforward way. Among other things students can tell a writer that an argument is not well introduced - they cannot see the relevance, or point, of it; that a section of the text is badly written - they are confused and cannot understand the points being made; that the examples given are poor ones, unilluminating or sending them off at a tangent; that an exercise seems pointless, takes too long or intrudes on the text and actually gets in the way of understanding; that points seem to have been explained in the wrong order; that they are bored stiff; that they are frustrated because the pace is far too fast, the facts raining down on them - they do not know what to do with them all; that they are infuriated because the tone is patronising. They can also tell a writer that the text is clear, lively, interesting and illuminating; that the exercise is appropriate - hard work but worth it; that the diagrams or illustrations are really helpful; that

the subject is fascinating. Though there are others who can tell a writer all these things - colleagues such as other writers, tutors, IET people - I suggest that they would be hard put to it to tell them in anything like the same detail or in a form that writers could as easily translate into action. Nor would their comments carry the same conviction.

In the introduction to this paper, and as its rationale, I claimed that writers' attitudes to course development, and to developmental testing in particular, are crucial. This is because any attempt to 'improve' a course, whatever form it takes, depends for its success on the writers' preparedness to accept the criticisms made, and to effect changes to their work. In saying this, I do not mean to charactise writers as stubborn or wayward creatures. Rather, I argued earlier that because writing is so revealing of the person, so effortful and time-consuming, writers understandably and rightly require to be convinced that someone else's view of what they have produced should be preferred to their own. It seems to me that students' views carry such conviction for a number of reasons. As I argued earlier, they are able by means of developmental testing to communicate their difficulties, ideas for improvement, and their more encouraging remarks at the right time, in a detailed and constructive way and, as I hope I have just demonstrated, over a very broad front. Moreover, of all those in a position to comment (leaving aside other subject specialists who tend to do so from an academic point of view) students carry the greatest conviction because what is written is intended for them - if it does not communicate itself to them it is nothing at all.

This brings me full circle. If writers do find students' views convincing and if, as I hope I have demonstrated, by means of developmental testing those views can be communicated over a broad front and with the kind of personal force that enables writers to form realistic and strong mental images of the students for whom they are writing, then it seems to me inevitable that this understanding will be reflected in the way writers approach the task of constructing courses thereafter. While I argued earlier that it is a mistake to expect students directly to contribute to discussions about course structure I did not mean to suggest that, consequently, writers are unable to glean anything of much relevance from the students' comments or from the impressions they form of the students on this basis. On the contrary, I think that those writers who said they had learned 'general' lessons from the experience of developmental testing were claiming just this. I think what they might have said was that as a result of getting to know even a few students intimately (that is coming to understand, as a result of what such students said and how they said it, what kinds of things they found difficult, foreign, time-consuming, off-putting, unexciting, insulting; and why; and how the students would have preferred things to be done) they had gained the kind of understanding that would enable them to go about (i.e. structure) things differently next time. I would add that this kind of insight (based on generalisation from the experiences of real students trying to get to grips with real courses) seems to provide a sounder basis from which to approach problems of course structure than does any amount of 'theorising' on the subject. It is in this sense (indirectly, via the perceptions of course writers), that I believe developmental testing can in the last analysis be said to affect course structure.

Finally, and more tentatively, I suggest that developmental testing can help writers in another way. It may be obvious that writers, even when members of Open University course teams (and especially in the attempt to write for 'distant' students?) are, while writing, in an essentially isolated position. What we have not so far considered is whether the impact of developmental testing can be explained in part by the fact that at this time it brings writers, students and evaluators together - that it makes writing more of a joint enterprise and imbues the whole process with the kind of collaborative spirit that should characterise teaching and learning in any setting.

Notes

1. It might be argued that in full-project courses students do just this. But such courses seem to me to compromise: students do have more choice about what they study but the parameters are normally laid down by the course team and the chosen topics vetted. Also the way students work, and the pacing of the project, is determined by the course team and monitored by tutors. It is worth noting that such courses are usually at fourth level.

2. As does the paper by Euan Henderson 'A241: A "better" course for students?' (pp. 5-9).

3. As does the paper by Barbara Mayor 'A292: What can we learn from developmental testing?' (pp. 10-15).

4. 'I can say in general that the feedback provided by the developmental testing of [my unit] of A101 was useful in many ways, and I should not wish to restrict this usefulness only to the question of "how to improve that specific unit".'

5. 'In general, developmental testing goaded me into thinking more carefully about what and how I was teaching ... with developmental testing, since each section of the unit is treated in some detail, there is no relaxing one's efforts ... This added awareness of what one is doing can be applied to any other teaching project.'

6. 'Most of all I was guided by the very careful and valuable advice which I received from [the evaluator] ... In addition to sending me her own detailed summaries and suggestions, she passed on to me the actual feedback forms returned by the students, with all their written comments.'

7. 'The part that caused the students most trouble was the long selection of definitions of religion, and the time-consuming exercises which they were asked to carry out using those definitions. Following [the evaluator's] advice I have slashed the definitions to a minimum, and have greatly simplified and reduced the exercises. All the way through I have been able to make extensive cuts and simplifications of the text.'

8. 'As you will see from the revised version of Section 4 of [the unit], your report on developmental testing was very useful. I have made a number of minor revisions and two major ones.'

9. 'I have revised the <u>discussion</u> (but not the question) on the Eskimo poem. Actually, I thought the students' answers on this more "acceptable" than you did. But I have <u>used</u> one or two of the answers to add to and at the same time change the discussion.'

10. Open-ended questions are also always included so that the <u>whole</u> unit is monitored for potential problems, not only those parts, or aspects of it, which writers or evaluators suspect might be difficult or open to misinterpretation.

5. U201: RESOURCE ASPECTS OF PRE-PRESENTATION COURSE EVALUATION

Paul Lefrere

Evaluation by Consultants

Speaking generally, this type of evaluation is similar in many ways to the (unpaid) conventional 'commenting on each other's drafts' which course unit authors are used to. This makes it prone to the same problems: it may not identify all the difficulties subsequently encountered by students when the course is presented in its printed form, and some of the potential weaknesses identified in advance of presentation may well not matter to the eventual students.

On most courses, consultants appear to have been used for two main purposes:

1. To provide written and verbal comments on the content and probable teaching effectiveness of course material (the current fee for this being £10 per unit or £320 per full credit per consultant);

2. To re-work any material considered by the consultant to be deficient (this more involved task commands a fee of £20 per unit).

The U201 Rising course team felt that either purpose was within their own capabilities, so decided instead to ask four part-time tutors to provide written and verbal comments on a rather different aspect of their course, its 'tutorability'.[1] A fee of £150 per tutor-consultant was therefore negotiated for reading all the first full drafts, commenting primarily on the difficulties which the tutor-consultants envisaged would be faced by tutors on the course, and secondarily on how students might react. (Detailed comments on the latter were being provided concurrently by a team of 'student consultants', as described later in this paper.)

As with the more conventional use of consultants, a detailed written brief needed to be prepared - in this case, by the Course Team Chairman and myself (the IET member of the course team). The total hidden costs of this exercise cannot be quantified exactly, but were similar to those in other course teams: course team time in liaising with consultants and in analysing their written feedback; making revisions; and providing each consultant with copies of unit drafts. In the case of U201 the consultants were expected to work to the same schedule for commenting as central course team members, and to comment on the same standard, unedited, reprographed material as was circulated to other members. Clearly, any deviation from these latter practices must involve additional costs.

My feeling, looking back, was that the use of tutors on U201 was justifiable, but not strictly necessary expenditure. However, it not only helped allay some of the concern of the Briefing and Training Sub-committee over the demands which this - the first U-course - would place on tutors, but also proved of some rather limited usefulness in revising the first draft material. However, a greater dependence on tutors (as in more involved and more expensive schemes) would, in my opinion, have been far less justifiable.

Developmental Testing for Payment

In addition to using tutor-consultants, the U201 course team also used ten Open University students as 'student-consultants', who were each paid £20 to give their reactions to the general pace, content, style, and organisation of the first full drafts, but not to work through (i.e. study) them.[2] There were two main reasons for limiting their task to giving reactions to 'presentational' aspects.

1. The breadth of this, the first U-course, meant that initial reactions to each new topic might matter more than on other courses, perhaps even affecting the dropout.

2. It was felt impossible to justify the greater amount of effort involved (and hence resources consumed) by any more ambitious scheme to evaluate these, our first full drafts. By way of illustration, the more usual testing-for-payment requires students to attempt self-assessment questions and other activities, to answer specific and detailed feedback questions, and to complete assignments, all for a fee of £10 per unit (or £320 per student for a full credit). To be effective, such schemes require the questions posed to be tailored to the units, in minute detail. This not only involves those course teams in a considerable amount of preparation, but generally requires the services of a 'feedback analyst' (sometimes specially employed) and the adherence of the team to a special handover schedule, sometimes modified to allow special student versions of the units to be produced, possibly with some pre-editing.

In the event, the quantity and type of information acquired from student-consultants by the U201 team seemed to represent a reasonable return for the costs involved, and was I think an important factor in their decision to change the authorship of some units and the emphasis of others. It is perhaps interesting to note that apart from the hidden costs (under the same headings as for the U201 tutor-consultant scheme), the cost of this exercise was £200 in fees, coincidentally the present (1980) maximum sum allocated to any course for developmental testing for payment. (I also note in passing that such a sum has been used by several other course teams for the purposes of testing selected aspects of their courses - for example, components of home experiment kits, detailed work on single units or blocks of several units, and less detailed assessment of the printed material.)

Developmental Testing for Credit[3]

In general, this method of evaluation is directed more towards finding out how effectively students really learn from course material than with establishing how they might react to it, or whether its content is correct. Accordingly, in the by now established form of testing for credit, 25-40 student testers can receive a course credit for successfully studying second drafts of course units, together with as many other components - including set books, assignments, examination and tuition - as are needed to make the course credit-worthy. Their course fees will also be reimbursed if they additionally provide a substantial amount of data on their reactions to the material, their study time, and so on. (Suggestions for revisions are based on inferences drawn from this data.)

This kind of exercise may not only involve the course team and other parts of the University in much extra effort, perhaps to a tighter schedule than normal, but can easily cost several thousand pounds, both in direct and hidden costs (Table 5.1). The hidden costs are complex in nature, with a large 'opportunity cost' component: the staff involved might have been doing something else with their time if they had not been involved in developmental testing, but one cannot know how important this 'something else' would have been to the University. It is therefore usual to consider only the direct costs, in which case it is hard to disagree with the assertion that testing for credit leads to improved teaching and at a cost which is small relative to the total cost of producing the course or remaking it earlier than would otherwise be necessary.

Table 5.1

Direct costs	
Tutor fee (contact hours/TMA marking/expenses)	£1000
Repayment of fees to students	£1595
Examiners'/markers' fees	£ 380
Audio cassettes	£ 14
Postage	£ 200
Reprographics	£ 850
Clerical assistant to collate and process feedback	£1320
TOTAL	£5359

Hidden costs	
Revision of artwork after developmental testing	
Editorial time prior to developmental testing	
IET time for devising feedback questions and preparing reports to the course team	
Secretarial time for typing feedback questions and revised units	
Course co-ordinator/course team chairman time for administering developmental testing programme	
Author time for making revisions[4]	
BBC producer time for editing after developmental testing	
University administration time	
Student Computing Service programmer time for making revisions after developmental testing	
TOTAL	£5000 - £10000

Although no-one on the U201 core course team had any direct experience of testing for credit, this was the only method of pre-presentation evaluation that seemed likely to provide us with the required quality and quantity of feedback on our second full drafts. When we decided to go ahead with it, we

looked in detail before we started at the testing carried out on A241, A101, etc., and found several ways of simplifying the most time-consuming parts (questionnaire writing, feedback analysis, etc.). This reduced the final costs of the exercise substantially.

The 'return' for this expenditure is also difficult to quantify. Some students failed to complete the requirements leading to fee-reimbursement, saving £275. There was a saving in copyright fees of about £300, since some articles which were planned to be included in the course were omitted or edited below 400 words as a result of feedback on developmental testing. Testing also resulted in some reduction in the number of printed pages required and in some reduction from the standard figure of 33% of print costs for printer's corrections. In addition, these reductions must have led to savings in editorial and studio time, but none of these items can be even approximately quantified. Above all, developmental testing has resulted in improvements in the quality of U201 and the experience will probably also influence the quality of subsequent courses to which U201 authors contribute: these also, however, are unquantifiable. The usual objective measures are (a) lower dropout (perhaps attributable, if it occurs on U201, to better tutor briefing by the tutor involved in the developmental testing), (b) higher demand for the course, (c) higher pass rate, and (d) greater potential for extending the life of the course.

One further return from developmental testing for credit must be mentioned. The University is largely financed by a block grant from the Treasury. One way of calculating the subsidy enjoyed by each successful student in a particular year is to divide the block grant by the total number of credits gained in that year. In 1979, the year in which U201 underwent developmental testing, the block grant was £34,526,000 and 40,801 full credits were gained. This gives a subsidy of £846 per student credit. A 'by-product' of the testing of U201 was that 29 student testers received a full credit, benefitting the University to the tune of £24,534 if 100% of the block grant is viewed as being support for undergraduate teaching activities. Even if only 50% of the block grant is regarded as being applied to undergraduate teaching, then this 'return' on the developmental testing for credit of U201 virtually balances the total of direct and hidden costs given in Table 5.1.

Subjectively, I feel testing for credit can lead to an increase in the effectiveness of course teams: for example, authors and IET staff acquire a more realistic perception of student capabilities, and a greater sensitivity to student needs. While I cannot attempt to put a value on effectiveness or increased morale, the U201 team generally feel that their units improved markedly after developmental testing. The feeling was that what was required of them in the way of revisions was very clear, partly due to the very detailed nature of the student feedback and the specificity of the recommendations, and partly because of the many helpful suggestions students made for additions or modifications to difficult material. These often helped overcome particular problems (for example, lack of clarity, ambiguity, inconsistency, and inaccurate cross-referencing), but were felt also to be valuable teaching experience, otherwise difficult to come by (even at summer school) in an institution like ours.

To summarise, testing a course by any method is to some extent an act of faith, since any improvements may be unquantifiable. While evaluation data from testing for credit schemes may be more reliable, and acted upon more often than in other schemes, it is also very time-consuming. Testing for credit should therefore not be undertaken lightly, or without good cause. However, had resources only been available for <u>one</u> form of evaluating U201, I am now convinced that testing for credit would have been chosen.[5]

Notes

1. This evaluation was conducted during 1978, two years before presentation.

2. This developmental testing exercise was also conducted during 1978, in parallel with the evaluation by tutor-consultants.

3. As has been usual in Open University practice, developmental testing for credit was conducted in 1979 - the pre-presentation year.

4. One of the more important resource costs of a developmental testing programme is the extra (or marginal) time and effort incurred by the authors of units over that which would otherwise have been expended. It is sometimes assumed, by those that have not been directly involved, that developmental testing for credit takes up an extra year of course team time. George Lang has attempted to show how such marginal costs might be calculated in principle, as follows.

Developmental testing makes its biggest impact on course team resources between the appearance of the second draft (D2) and handover of the final draft (D3), so we need to ask how much longer the D2-D3 process will take and how this affects the time taken by authors of units. In reality there is no such thing as 'normal' course production, with or without developmental testing, but for our purposes it will be assumed that D2s are produced on a regular basis at staggered intervals of y_1 days, and that a set period is allowed on each unit for course team reading and for the author to undertake revision and write D3, totalling x_1 days per unit. For a course with N units, the total time (T_1) to process the course, from the appearance of D2 of the first unit to handover of the last unit, is given by:

$$T_1 = x_1 + (N - 1)y_1.$$

With developmental testing for credit, the period x_1 may be longer (x_D) because the unit author cannot start revision until the results of developmental testing are known. Also, y_1 will be larger (y_D) if student study time per unit is greater than the handover stagger when a course is not tested in this way. However, given study times of 14 days for a half credit and 7 days for a full credit, it is unlikely that $y_D > y_1$. We will assume $y_D = y_1$.

The total time to process a course from D2 to D3 including developmental testing (T_D) is given by:

$$T_D = x_D + (N - 1)y_D.$$

The extra time required is therefore:

$$T_D - T_1 = x_D - x_1 + (N - 1)(y_D - y_1).$$

If $y_D = y_1$,

$$T_D - T_1 = x_D - x_1.$$

There are likely to be as many estimates of x_1 as there are members of the University, but let us take six weeks (42 days) as a reasonable average (two weeks for the course team to read D2 and four weeks for revision to D3). On the basis of my experience with a half credit course (D324 Business Economics), x_D involves the following:

Writing feedback questions and reprographic time	10 days
Mailing to students	10 days
Student studying time	14 days
Mailing feedback to University	10 days
Processing data and writing and reprographing report	14 days
Author's revision time	30 days
Estimate of x_D	88 days

So the extra time required to process a course subject to developmental testing for credit is 46 days - a far cry from the one year so often bandied about. For a full credit course, student reading time is only 7 days, so the estimate of x_D is only 39 days. Other possibilities exist. Two of the most likely are that $x_1 > x_D$ and/or that $y_1 > y_D$. In these cases developmental testing requires course teams to work faster than they normally would. As is implied in the discussion of Table 5.1, this may, of course, imply positive or negative marginal costs.

5. Jack Dowie, the Course Team Chairman of U201, adds the following comments.

I was merely convinced that developmental testing for credit was highly desirable before the U201 experience. After it (and exposure to other Open University courses' history) I am increasingly disposed to the view that it is scandalous that it is not normal practice - and becoming even less normal.

True cost-accounting of developmental testing is fraught with hazard. It is clear to me that no one knows what the costs of any course activity are under a systematic set of assumptions about what is fixed and what is variable in the University. And of course there is a variety of alternative sets of assumptions which could be made. But unless one is disposed to put a number on the benefits one is not entitled to argue on the basis of numerical costs. There is precious little logic in asking how much it is worth spending to improve a product from quality x to quality 1.1x (i.e. by 10%) if one is not willing to ask how much it is worth spending to achieve x. My personal impression is that the rate of return on developmental testing expenditure is much higher than the average return on the University's budget. Others have other hunches - and other sets of cost assumptions which reflect, naturally enough, their particular place in the organisation .

The other point I would like to make is that bringing part-time tutors in at first or second draft stage is not effective - because it is too late. They ought to be brought fully into the earlier stages of the production process. The most striking feature of the U201 developmental testing year was that both tutor and students found it highly rewarding compared with other courses. From their descriptions of what they experienced, I infer that it was simply what all university students (and teachers) should be experiencing.

6. S337: A NOTE ON THE DEVELOPMENTAL TESTING OF A SCIENCE COURSE

Barbara Hodgson

There are some fairly obvious differences between courses in the experimental sciences and those courses discussed in the previous sections of this paper. The text of a science course often relies heavily on graphical presentations of ideas and data and also on mathematical representations of concepts. Experimental techniques are taught through home experiment kits, project work and experimentally based summer schools. Particularly at third level, science courses may have several essential prerequisite courses, a situation which constrains both teacher and student. Television is usually considered an essential element of the course: it is used, for instance, to take students to otherwise inaccessible experimental or field locations or to explain and reinforce three and four dimensional concepts. Students are warned that inability to view television programmes may put them at considerable disadvantage. The material in science courses lends itself more readily than in some other disciplines to assessment through computer-marked assignments and most science courses use CMAs as a major component of their assessment procedures. Many of these course elements were developmentally tested recently in a third level Earth Science course.

The Course

S333 Earth Science Topics and Methods, a third level half-credit course, was first presented in 1976. The course deals, in case study form, with advanced topics in earth sciences and it soon became recognised that the course was extremely demanding and considerably overloaded. To alleviate this situation it was decided to add some extra case studies and then reorganise the material into two separate, more realistic, half-credit courses (S335 and S336).

Obviously these new courses would be excluded combinations with S333, so that it was necessary to arrange that those students who already had credit in S333 would not be penalised by this situation. To this end it was agreed that we would present the new components of S335 and S336 as a 'special' course which would be open only to students with credit in S333 and which, for such students, would count as a half-credit.

Because of the problems with the original course (S333) it was felt advisable to undertake developmental testing of the new components in the form of this 'special' course, S337 Earth Science Topics and Methods Part II.

Developmental Testing Arrangements

The population of student testers for S337 was well defined and recruitment was easy. Because of the very limited life of S337 (testing year + 2 years), it was advantageous to allow a reasonable number of students to study the course in its developmental testing year to help ensure that everyone qualified to take the course had an opportunity to do so. We were allowed to recruit eighty testers.

This large number of testers meant that we were able to make use of the standard item analysis procedure in the testing of CMAs. However we felt that we might have an excessive amount of detailed data to analyse so it was agreed that each student would provide limited feedback throughout the course and detailed feedback on just one case study. Students were then given a remission of half the course fee on completion of testing. These arrangements kept both the cost and student workload within reason.

The potential testers were too widespread geographically to be based on any one study centre. To make viable tutorial groups possible we aimed to recruit groups of twenty students in each of four regions. The tutors were all experienced in tutoring S333.

Because arrangements had already been made to include two courses rather than one in the broadcast transmission schedule, slots were available for television and radio programmes for S337 to be broadcast during the developmental testing year. Consequently television and radio were tested under normal circumstances. Otherwise the developmental testing procedure followed during the course was much as described elsewhere in this paper.

Some Outcomes

As with other courses, the testing of S337 did not lead to major changes in the structure or content of the course. Many small changes were made including the expansion of an explanation of a complex three-dimensional concept, the inclusion of more worked and student examples and the rephrasing of arguments to allow students to be more involved in making and justifying decisions and choices based on evidence presented. One case study was reduced in length considerably when it was realised that the students felt its introduction overlapped the topics of a prerequisite course.

There were however two areas of the course which were considerably modified after developmental testing - the illustrations and the project essay.

It was intended that the illustrations should be presented in a 'final' drawn state and this was so for the first case study but later production difficulties meant that author's sketches and photocopies of other suitable diagrams had to suffice. All illustrations were in tones of black and white. It soon became obvious that, even on the 'final' drawings, the density of information on many of the illustrations, especially the geological maps, was just too great to be easily distinguished and understood when the illustrations were only in black and white. This gave support to the course team's bid for resources for two-colour printing. Many small errors and ommissions - the sort of things which, though small, tend to render illustrations useless for their intended purposes - were corrected. In one case study, which was rather too copiously illustrated, the testers were very helpful in suggesting which illustrations might be dispensed with.

For their project the students were asked to study several research papers and to write a long essay on bias and reality in the way geologists present their findings. After seeing how the testers tackled this, their comments on it, and the length of time it took to complete, the project was considerably restructured. It is now done in two parts, involves less reading, and students are given more guidance about tackling it. It also runs concurrently with other parts of the course instead of being the first thing that students are required to tackle.

This course did not have a home experiment kit and the summer school was not tested as the students had already completed it while studying S333. We will have to wait until S271 Discovering Physics is tested in 1981 to assess the outcomes of the developmental testing of experimental summer schools and home experiment kits.

COURSE DEVELOPMENT GROUP PAPERS

CD1 NOTES ON THE EARLY STAGES OF COURSE PLANNING
by Phillipe Duchastel, Roger Harrison, Euan Henderson, Barbara Hodgson, Adrian Kirkwood, Robert Zimmer

CD2 HOW TO USE CONSULTANTS SUCCESSFULLY
by Judith Riley

CD3 DISCUSSING AND EVALUATING DRAFTS
by Judith Riley

CD4 ASSESSMENT
by Judith Riley

CD5 INTRODUCING NEW FACULTY MEMBERS TO COURSE PRODUCTION
by Phillipe Duchastel and Roger Harrison

CD6 BRIDGING THE GAP BETWEEN TEACHING AND LEARNING AT A DISTANCE
by Michael Nathenson

CD7 COLLECTING FEEDBACK DURING COURSE PRESENTATION
by Fred Lockwood

CD8 DESIGNING INSTRUCTIONAL MATERIALS: GUESSWORK OR FACTS?
by Stephen Brown and Michael Nathenson